THREE RELATED PROGRAMS FOR RETRIEVAL, PROGRAMMED TEACHING, AND THE TAKING OF NOTES

Ted Nelson
20 Mt. Auburn St.
Cambridge 38, Mass.
March, 1962.

THREE RELATED PROGRAMS FOR RETRIEVAL, PROGRAMMED TEACHING, AND THE TAKING OF NOTES

This is a hurried sketch of three closely related types of programs that I think should be written, in at least crude form, for high-speed machines with tape drives and some sort of on-line output and input. For logical simplicity I will assume the use of a simple machine (e.g., the Packard Bell 250) with a typewriter for both output and input and a couple of magnetic tape drives. However, it should be evident that economies of scale for some of the functions mentioned below will not appear until larger machines can service many stations.

1. "N-category" revisable cataloguing systems, especially for information retrieval.

These systems use machines for sorting, but according to principles that are conceptually different from sorting in the usual sense. By n-category tagging I mean that an item (such as the title of an article) can be coded as being in or out of any number of categories, and a new category can be added at any time. A record is kept of all the categories an item is definitely in or out of. However, a decision does not have to be made as to whether a given item is in a given category.

New items may, of course, be filed and checked by a coder as being in or out of some range of the categories in storage. The categories may have titles, but the titles can be changed on examination of the items in the categories. Moreover, new 'purified' categories may be made from old ones. That is, new categories are set up which contain only some of the contents of the previous ones.

Items can be recoded from time to time.

Searches can be requested; the machine will then present a list or a count of the contents of a category, or of the things conjunctly or disjointly in several categories, or all items being in at least a certain number of categories from a given list, etc.

Inventory provisions: The number of things on file and various counts of the distributionof their category tags, can be made automatically. Moreover, a history of the file, including inputs and changes, permits the return to any previous state which was better.

In some systems it may be desirable to store entire books and articles on magnetic tape, and feed them out to the user as required. While this is not critical to the philosophy of categories here, it is assumed later in the paper.

Discussion.

The paramount fact to me is that categories change in their meaning. Partly because there are subtle shifts in the semantics of category titles, partly because the subject matter changes in time or comes to be looked at differently, partly because the items within a category may turn out to have autonomous properties of their own that were not anticipated, it is a mistake to expect categories, especially important ones, to be stable.

In this respect I take the "keyword" principle, as currently used in IR theory, to be spurious. The fact that words are selected as 'objective' criteria of the categories to sort things in is misleading. "It usually takes a highly-skilled person to select keywords in a document..." (IBM, "An Introduction to Information Retrieval," (p.10). That is, the keywords are only part of the criteria by which they are assigned to categories, and their objectivity provides only hints for their categorization (or in the extreme case, excuses). Although many categories remain stable and recognizable by the same keywords, a crucial fraction are always in conceptual change. Are not these the ones for which retrieval is the biggest problem?

2. Easy-load-format n-pathway teaching programs.

This program operates on materials already on tape, i.e., books and articles which are in the machine already. The Proprietor of the program selects sections of the material which he wants shown to students, and decides on possible 'pathways' through the material. That is, a student may proceed in linear fashion, seeing one thing after another, not having to respond.

The Proprietor may also add interpolated items--for instance, questions and connective material--that were not previously on file. He may also put in branchings--items which ask the student a question, or request him to state his preference. The student's response will determine, according to the previous instructions of the Proprietor, what materials he will see subsequently.

Note that the usual linear sequence is by no means the only one possible. In the linear sequence the student sees item A, gives a response, and then sees A1, A2, ...An, or (in the case of a correct answer) B. He will never see C unless he sees B first. In this general program such a convention is strictly the option of the Proprietor. By "n pathways" I mean it would be possible to add conditional pathways indefinitely.

Discussion.

Programmed teaching today is focussed on finding highly efficient items to teach certain very specific things to a target population. Whether this is the proper approach depends on some very important assumptions, for instance, the costs of using inefficient items (in terms of time wasted and mis-learning), in comparison with the enormous cost of weeding out the inefficient items.

The alternative described here--the easy-load, multiple-pathway program--has these advantages: it can be quickly programmed ad hoc in a particular subject for a small population, e.g., a small class taking a specialized course; such a program need be no more complicated than a reading list, but indefinitely more flexible because it can be specially branched for a student's misunderstandings and interests. Furthermore, it can be easily updated, improved or revised as its proprietor sees the need.

Low efficiency of first-attempt loading (compared to professional high-cost teaching programs) is offset by the great diversity, variety, and cheapness of them, and with their vast adaptation to the students.

3. Programs for taking notes on written material.

Supposing that someone is reading material presented him by the computer, as in (1), he may type in notes for his own future benefits, which are stored on his own tape, and request the storage of certain parts of the original in his own notes--for his own future use as quotations in his own writing, or to remind him of the contents of what he read.

The materials may then be arranged by the user in "n categories" (by system 1), or in alternative sets of sequences that he may later go back through (by system 2).

Discussion. This would be of use to a student at any level, as an economical means of summarizing what he has read, and condensing it for review and emphasis. Advanced researchers could quickly excerpt key passages and quotations.

GENERAL REMARKS

I see information retrieval (which is really the category problem), programmed learning and the organization of notes as three elements of a united whole.

Information retrieval at most levels involves showing somebody something he doesn't know, even if he is an advanced professional in his field. There is no reason why the authors of professional articles might not prepare their materials in multiple-pathway form, to depend on the interests and knowledge of the reader--much as they now prepare summaries without complaint. Or this could be done by editors.

The growing capacities of machine storage permit repository libraries to be founded where researchers may place fringe materials in one way or another unsuitable for journal publication. Similarly, they would also permit the storage of unusual teaching programs--peculiar materials, or peculiar treatments, or aimed at peculiar populations--for instance, Hopi Indians wishing to study

nuclear physics (for which, it is rumored, they have a special aptitude.) In previous times such material was far more perishable, and perhaps would never be attempted because it could never be found again. Today it seems increasingly possible that such things could be found again.

This combination of interwoven systems would have the advantage that one running it could be programmed immediately for use in any field of knowledge without fear of obsolescence because of the updating features. (The problems of charts, illustrations and mathematical symbols are ignored here.)

Furthermore, the possible variety of such instruction programs would probably have a tremendous impact on certain problems in education where misunderstandings, impatience and anxieties are high, such as the teaching of mathematics to people who have shied away from it. As a rule teachers of mathematics cannot fathom the impact that the failure to understand will have on a student, say, of literature, who has built so much of his career on his ability to achieve immediate insight. Such programs, put together experimentally by people here and there and elaborated by others, would find gradually what approaches work.

Hence the emphasis here is evolutionary. The idea is to get the programs working and watch them improve, rather than do intense research for a year before putting one finished program on the market.

A combination of them--which seems 'ultimate,' at least in 1962--would permit the automation of libraries and the whole process of study, if not of teaching. I cannot help thinking that this is a "great leap" which the underdeveloped countries could make.

Ted Nelson
20 Mt. Auburn St.
Cambridge 38, Mass.
March, 1962

Three Related Programs for Retrieval, Programed Teaching, and the Taking of Notes.

This is ~~just a brief~~ a hurried sketch of three closely related types of programs that I think should be written, in at least crude form, for high-speed machines with tape drives and some sort of ~~mux~~ on-line output and input. For logical simplicity I will assume the use of a ~~Packard Bell 250~~ simple machine (e.g., the Packard Bell 250) with ~~two tape machines and a Flexowriter~~ a typewriter for both output and input. However, it should be evident that ~~for~~ economies of scale for some of the functions mentioned below will not appear until larger machines can service many stations.

1. Preliminary points. ~~[illegible]~~ There are several assumptions which underlie my thinking here. First,

1. Information retrieval. I think there is something a trifle spurious about the "objectivity" of keyword systems

1. "N-category" revisable cataloguing systems, especially for ~~one filing and~~ information retrieval. ¶ These systems use machines for sorting, but according to principles that are conceptually different from sorting in the usual sense. (such as the title of an article)

No ¶ By n-category tagging I mean that an item can be coded as being in or out of any number of categories, and a new category may be added at any time. ~~and a new cates~~ ~~and a new category may be added at any time.~~ ~~¶ A~~ record is kept of all the categories an item is definitely in or out of. However, a decision does not have to be made on a given category as to whether a given item is in. New items may, of course, be ~~filed~~ filed and ~~the cat~~ checked as being in or out of some range by a coder of the categories in storage. No ¶

~~[illegible]~~

~~[illegible]~~

The categories may have titles, but the titles can be changed on examination of the items in the categories. Moreover, new 'purified' categories may be made from old ones. ~~by removing~~ ~~sort~~ That is, ~~setting up~~ only some new categories are set up, which contain ~~certain parts~~ of the previous ones. ~~While there~~ of the contents ~~may be more~~ There is no reason for categories to be destroyed. Rather, more importantly, ~~these~~

Items can be re-recoded from time to time.

Searches can be requested; the machine will then present a list of the contents of a category, or a count of the contents of several categories, or of the things conjunctively or disjunctively being in at least a certain number of categories, or all items in two categories, etc, from a given list, etc.

Inventory provisions: the number of things on file and various counts of the distribution of their category tags, can be made automatically. Moreover a history of the file, including inputs and changes, permit the return to any previous state which was better.

→ Discussion. The paramount fact to me is that categories change in their meaning. Partly because there are subtle shifts in the semantics of category titles, partly because the subject matter changes in time or comes to be looked at differently, partly because the items within a category may turn out to have autonomous properties of their own that were not anticipated, it is a mistake to expect categories, especially important ones, to be stable.

In this respect I take the "keyword" principle, as currently used in IR theory, to be spurious. ~~The use of keywords for automatic tagging is admitted (e.g., the Information Retrieval Electronic Storehouse manual put out by IBM)~~ The fact that words are selected as 'objective' criteria of the categories to sort things in is misleading.

In some systems it may be desirable to store entire books and articles, as this is not critical to the philosophy of categories here. It is assumed later on that the user is required. While

No 4

"It usually takes a highly skilled person to select keywords in a document..."
(IBM, "An Introduction to Information Retrieval," p. 10.) Although
many categories remain stable and recognizable by the same keywords,
a crucial fraction are always in conceptual change. Are not these
the ones for which retrieval is the biggest problem?

That is, the keywords are only part of the criteria by which they are assigned to categories, and their 'objectivity' provides only hints for their categorization (or in the extreme case, excuses.)

2. Easy-load-format, teaching programs.

This program operates on materials already on tape, i.e., books and articles which are in the machine already. The Proprietor of the program selects sections of the material which he wants shown to students, and decides on possible 'pathways' through the material. That is, a student may proceed in linear fashion, seeing one thing after another, not having to respond. The Proprietor may also add interpolated items — for instance, questions and connective material — that were not previously on file. He may also put in branching items which ask the student a question or request him to state his preference. The student's response will determine, according to the previous instructions of the Proprietor, what materials he will see subsequently.

Note that the usual linear sequence is by no means the only one possible. In the linear sequence the student sees item A, gives a response, and then sees A1, A2 ... An, or B (in the case of a correct answer B). He will never see B unless he sees B first. In this general program such a convention is strictly the option of the Proprietor. By "pathways" I mean it would be possible to [illegible] teaching indefinitely.

DISCUSSION. Programmed teaching today seems to be focused on finding highly efficient items to teach certain very specific things to a target population. Whether this is the proper approach depends on some very important assumptions, for instance, the cost of using inefficient items in terms of time wasted and mis-learning, versus the cost of weeding out the inefficient items.

The alternative described here has those advantages: it can be the easy-load, multiple-pathway program —

quickly programmed ad hoc in a particular subject for a small population, ~~dealing with a particular subject~~, e.g. a small class taking a specialized course; such a program need be no more complicated than a reading list, but indefinitely more flexible because it can be specially branched for a student's misunderstandings and interests. Furthermore, it can be easily updated, improved or revised as ~~the~~ its proprietor sees the need.

~~The intricacy of reading lists~~

Low efficiency at first-attempt loading compared to professional ~~Direct~~ high-cost teaching programs) is offset by the great diversity, variety and cheapness of them, and ~~together~~ with their vast adaptation to the students, ~~would give them great importance.~~

3. Programs for taking notes on written material

Supposing that someone is reading material presented him by the computer, as in (1), he may type in notes for his own future benefit, which are stored on his own tape, and request the storage of certain parts of the original in his own notes — for his own future use as quotations in his own writing, or to remind him of the contents of what he read.

The materials may then be arranged by the user in "categories" (by system 1), or in alternative sets of sequences that he may later go back through (by system 2.)

Discussion. This would be of use to students at any levels, as an economical means of summarizing what he has read, and condensing it for review and emphasis. Advanced researchers could quickly excerpt key passages and quotations.

GENERAL REMARKS. ~~these three forms of~~

A combination of them — which seems 'ultimate,' at least in 1962 — would permit the automation of libraries and the whole process of study, if not of teaching. I cannot help thinking that this is a "great leap" which the underdeveloped countries could make ~~step~~

~~Furthermore, if~~

This combination of interwoven systems would have the advantage that once running, it could be programmed immediately for use in any field of knowledge without fear of obsolescence because of the updating features. (The problems of charts, illustrations and mathematical symbols are ignored here.)

~~Most important, I want to understand why~~ I see ~~see information retrieval~~ information retrieval (which is really the category problem,) programed learning and the organization of notes ~~and information~~ as three elements of a united whole.

Information retrieval at ~~the level of profes~~ most levels involves ~~the showing of~~ showing somebody something he doesn't know, even if he is an advanced professional in his field. There is no reason why the authors of professional articles might not prepare their materials in multiple-pathway form, to depend on the interests and knowledge of the reader — much as they now prepare summaries without complaint. Or this could be done by editors.

GO ON & COME BACK LATER

21 The growing capacities of machine storage permit repository libraries to be founded where researchers may place fringe materials in one way or another unsuitable for journal publication. Similarly, they would also permit the storage of unusual teaching programs — peculiar materials, or peculiar treatments, or aimed at peculiar populations — for instance, Hopi Indians wishing to study nuclear physics (for which, it is rumored, they have a special aptitude.) In previous times such material was far more perishable and perhaps would never be attempted because it could never be found again. Today it seems increasingly possible that such things could be found again.

(3) Furthermore, the possible variety of such instruction programs would probably have a tremendous impact on certain problems in education where misunderstandings, impatience and anxieties are high, such as the teaching of mathematics to people who have shied away from it. As a role mathematicians cannot fathom the impact that the failure to understand will have on a student, say, of literature, who has built so much of his career on his ability to achieve immediate insight. Such programs, put together experimentally by people here and there and elaborated +

by others, would find gradually what approaches work.

Hence the emphasis here is evolutionary. The idea is to get the programs working and watch them improve, rather than do intense research for a year before putting one finished program on the market.

GO BACK TO (4)

To T. Nelson 3/8/62

5 pp. @ 75¢ 3.75

+3

15 pp @ 5¢ .75

$4.50

(retyped 2 pp. because illegible handwriting!!)

(add misplaced or no carat insertion mark)

bb